Using Expert-Derived Aesthetic Attributes to Help Users in Exploring Image Databases

Cormac Hampson[1], Meltem Gürel[1], Owen Conlan[1],

[1] Knowledge & Data Engineering Group,
Trinity College Dublin, Ireland
{cormac.hampson, gurelm, owen.conlan}@tcd.ie

Abstract. Image repositories often contain a large amount of metadata about their content. However many resources, such as photographs, have inherent aesthetic qualities that can be difficult to describe in a semantically consistent and usable manner, yet would be highly valuable for users in exploring large image repositories, such as Flickr. Automatically augmenting existing metadata with expert perspectives has the potential to give users a consistent aesthetic vocabulary to search and explore such repositories. SARA (Semantic Attribute Reconciliation Architecture) is a system that supports users to leverage domain expertise while searching for items in a metadata-rich domain. X2Photo is a tool built on SARA's functionality to enable image searching based on a picture's aesthetic characteristics and user-generated tags. This paper describes X2Photo in detail, the approach to augmenting visual media with expertise, and the evaluation results which reveal how semantically described aesthetics can support complementary search axes for image retrieval.

Keywords: Image Retrieval, Subjectivity, Expertise, Crowd Sourcing, Semantic Attributes, User exploration

1 Introduction

Visual media stored in image repositories typically contain a lot of descriptive and technical metadata, be it user-generated tags, or data extracted from low level features or from content analysis. However many resources, such as photographs, have inherent aesthetic qualities that can be hard to describe in a semantically consistent and usable manner. This can make it difficult for users to explore image repositories for relevant photographs from an aesthetic perspective. In this paper we consider aesthetic attributes as means of describing emotions associated with specific images e.g. sadness, excitement, joy etc. Typically users are reduced to keyword searching over an image's associated tags, which tend to focus more on the content of the image (what it is portraying) rather than its aesthetic. Moreover, due to its subjective nature, any tags describing an images aesthetic are likely not to be semantically consistent from one tagger to the next, adding to the difficulty of locating appropriate images. For example people are quite good at identifying and tagging elements of a photograph, i.e. it contains a dog on a beach, but not in capturing its aesthetic i.e. it is tranquil and cool.

Advances in digital photography technology and related internet storage services have given users unprecedented ability to capture photographs and make them available to a large audience. Many popular image search engines, such as Flickr[1], utilise the metatags associated with a photograph. Unfortunately, they may be inaccurate and misleading at times [1]. While these tags can be useful in defining the "content" of a photograph, their single dimensional nature can limit more refined searching. Thus, consistently exposing the aesthetics of an image as a criterion that may be searched upon presents a significant challenge.

Introducing the use of expert knowledge into the exploration of photograph collections can enable end-users to discover more accurate results, and can help guide them through the process by allowing them to leverage domain specific identifiers. This may not overcome the subjectivity of the terminology, but will at least provide a consistent reference point for the domain. Furthermore, the knowledge of domain experts can lead the end-users to find photographs through the expert's vocabulary which may not have been obvious to the end-user when initially defining the photograph they were seeking. This is because domain experts have clear perspectives when it comes to defining the key characteristics of a domain. For instance, there is an expert domain dedicated to the subject of wine. When analyzing a particular type of wine, different perceptions such as colour /clarity, bouquet and taste are used by experts of this domain to express their sensations through descriptive words or phrases. The wine-tasting terminology comprises of high-level subjective terms which are derived from low-level characteristics of a wine. For example, the term "bitter" typically refers to the tannin content of a wine, and the term "oily" is used for the combination of high glycerine and slightly low acid content [2].

This expert terminology creates a semantic space for wine tasting. The words used can differ among experts but are based on the same characteristics. Applying similar techniques to the photography domain can help define a more automated clear-cut aesthetic search environment. Even if you may not totally agree with the expert's terminology in such a subjective area as aesthetics, you will have a consistent view on the domain and will know what type of images to expect in the results. Moreover if users can personalise and tailor an experts domain view to their own, it can give a user even more control and flexibility when searching for photographs.

This paper examines to what extent the combination of tags and subjective expertise, can support end users in exploring visual media? Here we refer to tags as annotations, such as those that Flickr users have assigned to their digital photographs. The expert knowledge is automatically derived from non-textual low-level data contained within a digital photograph; specifically the hue, saturation and lightness of its dominant colours. It is proposed that when combined, these two features should enable exploration from a content perspective (achieved via the selection of tags) and from an aesthetic perspective (derived from the expert knowledge). To investigate this, an application called X2Photo that works in tandem with the SARA [3] (Semantic Attribute Reconciliation Architecture) middleware system has been built and is described within this paper. The remainder of this paper is organised as follows: Section two highlights some related work and problems in the field; Section three discusses the design and implementation of X2Photo, as well as how SARA's

[1] http://www.flickr.com

authoring tool SABer (Semantic Attribute Builder) was used by an expert in the photographic domain to describe aesthetics. Section four describes the evaluation of X2Photo and section five summarises the research discussed in this paper.

2 Related Work

Most of the current research in image retrieval is concerned with bridging the semantic gap. In essence, this is the gap between the low-level physical features of the image and the high level perception of what the image portrays. As Hare states [4], the representations one can compute from raw image data cannot be readily transformed to high-level representations of the semantics that the images convey. It is these semantics in which users typically prefer to articulate their queries.

Research activity in visual image retrieval increased following the adoption of Content-Based Image Retrieval (CBIR). CBIR is the method of retrieving images on the basis of automatically-derived features such as colour, texture and shape. These systems try to retrieve images that are similar to a specification or pattern (e.g. shape sketch, example image) a user defines. The automatic retrieval process within these systems suggest an advantage compared to keyword based search systems as there is no possibility of the necessary metadata not being present. However, the limitations of current content-based retrieval approaches and their incompatibility with searchers' queries are often pointed out [4, 5]. The major obstacle in CBIR approaches is the gap between visual feature representations and semantic concepts of images. In general, the problem with these algorithms is their dependency on visual similarity in judging semantic similarity [6]. Especially for photographs, it is very difficult to devise effective features that reflect their aesthetic characteristic. As semantic similarity is a highly subjective measure, it is not reasonable to rely on such algorithms, especially when the semantic space comprises of aesthetic values.

Image retrieval based on keyword features [7, 8] was mainly developed by the database management and information retrieval community. The typical query scenario in such image retrieval systems is Query By Keyword (QBK). In this process the semantics of images are represented by keywords, with query results being acceptable if the keyword annotations are accurate and complete. However, as the size of the image database gets larger, manual annotation cannot be regarded as a viable procedure to continue. Popular image search engines such as Google[2], Yahoo![3] and Bing[4] try to overcome this issue by extracting the keyword features surrounding an image on the Web. Although this method can find numerous results, the returned images are not entirely accurate since there is no guarantee that surrounding textual information relates directly to the image. Likewise, when trying to attach semantics to visual content, you have the problem of dealing with homonymy, where a single tag may have various meanings. Hence engines that retrieve images indexed through such methods can only be accurate within a certain limit [9].

Flickr is an online community platform that enables its users to upload, store and organise digital photos. Features that Flickr uses to strengthen its metadata are to

[2] http://www.google.com/imghp
[3] http://images.search.yahoo.com
[4] http://www.bing.com/images

allow users to group their photos into sets, and their sets into collections. Moreover users of Flickr can create and add photographs to special interest groups on any possible topic, improving the relevance of a photo's metadata. Within Flickr, users and their contacts form the backbone of photograph propagation. Research indicates that social browsing, i.e. finding photographs by browsing through the photograph streams of contacts, is one of the primary methods by which users find new images on Flickr [10]. This suggests that in such an environment users are likely to "follow" other users and that photograph enthusiasts welcome the idea of expert guided browsing.

Looking at the current approaches, it is apparent that bridging the semantic gap is still an open issue. Indexing based on surrounding textual information is highly unreliable, and textual annotations depend on the knowledge and expressiveness of individuals, which causes ambiguity. Retrieving images through this textual data often results in inaccurate and irrelevant clusters of images. Furthermore, current technologies that are based on low-level visual information do not allow users to search for images by higher-level semantics. The need to provide initial query images or to find images based on unintuitive low-level characteristics explains why these approaches haven't yet found a noticeable place in the commercial world. Regarding photography appreciation, both of these approaches, though acceptable for defining content, are inefficient in reflecting the aesthetic characteristics of images. Sinha and Jain [19] point out that content only is not enough in inferring the semantics of photographs and suggest fusing content and context to extract semantics; referred to as a contextual analysis. Enser [20] also suggest that it is necessary to utilise both the concept and the content of a photograph to improve the efficiency of image retrieval techniques, stating that hybrid image retrieval systems should be welcomed.

Being able to retrieve images from image repositories using high level semantics defined by experts may help these systems to realise their potential more. Likewise, when introducing subjective qualities such as aesthetics as search criteria, it would be useful to have systems that can support personalisation within the process. By combining this with a user interface that supports end-users to manipulate photograph collections in a personalisable and compelling way, the system would empower users in exploring and accessing large image repositories.

3 X2Photo

The previous section highlighted how image retrieval techniques that combine textual annotations or keyword search with low-level characteristics are being considered in order to bridge the semantic gap. X2Photo is an application designed to help tackle this problem and is described in detail in this section. SARA [3] (Semantic Attribute Reconciliation Architecture), the middleware that X2Photo is built on, is described briefly next.

3.1 SARA

SARA is a domain independent framework that allows for low level metadata to be aggregated into semantically meaningful characteristics that ordinary users can understand. These characteristics (called semantic attributes) are defined by experts,

and then leveraged by end users to help their exploration of a domain. The semantic attributes can be objective or subjective in nature, and SARA can support the tailoring of these characteristics to an end users perspective or context. By adding these semantically meaningful concepts to a space that didn't have them before, it supports end-users (via an appropriate client application) to employ expert knowledge to create high-level, semantically meaningful queries over multiple sources from a domain. Essentially, these semantic attributes can be seen as generalised rules for the domain and SARA acts as a semantic mediator between end-users and the raw data sources they seek to explore. Importantly, these semantic attributes can be generated by non-technical domain experts without the help of a knowledge engineer by using SARA's authoring tool SABer (Semantic Attribute Builder). This means that semantic attributes can be generated by experts from almost any domain.

SARA has already been successfully applied to a number of domains including music, films, digital humanities and publications. Hence, its support for subjective semantic attributes, based on aggregated low level data, meant it was an ideal system to help the exploration of image repositories from an aesthetic perspective. The main aim of X2Photo is to use the functionality offered by SARA to help users browse large image repositories with reference to the aesthetics of the photographs, as well as their content. Specifically it supports the retrieval of Flickr photographs using domain expertise in aesthetics, as well as user generated tags. In order for X2Photo to use SARA it needed an expert vocabulary to describe the aesthetics of digital photography. This vocabulary would then be leveraged by end-users to give them a consistent approach to browsing for images. The next section describes why a vocabulary based on colour psychology and colour theory was chosen for X2Photo.

3.2 Colour Theory and Colour Psychology as an Aesthetic Vocabulary

Within CBIR (Content Based Image Retrieval) approaches, colour has been seen as a key feature to characterise the content of digital content collections [11-13]. Common colour features include, colour-covariance matrices, colour histograms, colour moments, and colour coherence vectors. Even though these colour features are efficient in describing colours, they are not directly related to high-level semantics. Hence, one way to derive human perception through colours is to investigate the psychology of colour in art [14, 15]. Artists use colour to explore visual perception and to represent or evoke emotions. The psychological effects of colour, hue, saturation, and brightness have been studied to reveal having various effects on the viewer [16, 17].

Complementary to colour psychology is colour theory, which is a language that conceptually and perceptually describes the essentials of colour and their interactions [18]. Unlike colour psychology, colour theory doesn't describe responses that are unique to cultures or certain periods, but rather focuses on universal psychological responses to colour. An example would be the warmth or coolness of a colour, i.e. the temperature. Colours such as blue and green are cool colours and can be thought of as having calming effects. However, this effect can transform as the colour's luminosity changes, i.e. a bright open sky may be exciting. Likewise, cool colours on one end of the scale can be seen as cold, impersonal, and gloomy but on the other comforting and nurturing. In photography, colour theory is utilised to understand how certain colours

and their combinations create different moods in photographs. For instance, some colour combinations such as complementary colours appear striking and vibrant when in close proximity. Furthermore, as they get closer to the same saturation and lightness, the vibrant look will strengthen. On the other hand, colours that are close in the spectrum will usually appear more peaceful and calm.

X2Photo required an expert vocabulary based on raw low-level data of digital photographs. By extending the language that colour theory provides with subjective concepts from colour psychology, a consistent vocabulary was developed that helped the exploration of photographs from an aesthetic perspective. The user could adapt to the expert's perspective or find it open to questioning. However the important factor here was not the vocabulary, but rather providing a base that a subjective concept could be built upon and if necessary personalised.

3.3 Design

Based on the information in the previous section it was decided that for representing colour in digital images, the HSL (Hue, Saturations, Lightness) colour model would be the most efficient regarding this research's aims. HSL colour space describes perceptual colour relationships more accurately than RGB and is far more intuitive. Fortunately, this kind of metadata is commonly found in digital images or can be easily extracted from the photograph. HSL colour space is also more closely related to human visual perception. Another point deducted from colour theory was that the human eye is more sensitive to hue than saturation and lightness. Therefore hue should be processed with a finer quantisation.

The expert vocabulary created for X2Photo consisted of nine semantic attributes, each with three or four parameters that were encoded by the domain expert in SABer (see Table. 1). *Temperature* was one such semantic attribute for the photography domain created, and this was quantised into a number of different parameters ranging from *Warm* to *Cold*. Thus *Cool* was a single parameter of the semantic attribute *Temperature*. When *Temperature* was defined, the hue of the colour was taken into consideration as follows. Colour theory defines colours such as blue and green as cool colours with red and orange defined as warm. Hue is represented as an angle of the colour circle. So if it is divided into twelve equal intervals, on each 30°angle, the following colours result; red, red-yellow (orange), yellow, yellow-green, green, green-cyan, cyan, cyan-blue, blue, blue-magenta, magenta, and magenta-red. Red, yellow, green, cyan, blue, and magenta are regarded as the key colours with each having intermediate colours in between. Thus, the classification for Temperature into four parameters was constructed as follows, where H, S and L represent hue, saturation and lightness respectively:

$$\text{WARM} = \{(0 \le H < 75 \text{ and } 15 \le L \le 90) \text{ or } (H \ge 300 \text{ and } 65 \le L \le 90)\} \text{ and } (S \ge 25)$$
$$\text{SUBTLE} = (75 \le H < 120) \text{ and } (15 \le L \le 90) \text{ and } (S \ge 25)$$
$$\text{COOL} = (120 \le H < 210) \text{ and } (15 \le L \le 90) \text{ and } (S \ge 25)$$
$$\text{COLD} = (210 \le H < 300) \text{ and } (15 \le L \le 90) \text{ and } (S \ge 25)$$

If a photograph's colour space satisfies the third equation, it is considered as a *Cool* photograph. Table 1 lists the nine semantic attributes created for X2Photo, each with

three or four different parameters. These subjective semantic attributes were created in a similar way to the *Temperature* example described above, and were also joined by one objective semantic attribute named *has tag called* that allowed users to specify tags that the end images should have. All the semantic attributes created were listed in the X2Photo interface so that they could be joined together into a complex query by end users e.g. *Return all images that are Calm, Cool and Misty that has a tag called Boat or Fisherman*.

Table 1. The nine semantic attributes and their parameters created for X2Photo

1. Power	**2. Passion**	**3. Energy**	**4. Joy**	**5. Ease**
Vigourous	Passionate	Explosive	Frantic	Easeful
Powerful	Desirous	Exciting	Ecstatic	Content
Robust	Romantic	Energetic	Jolly	Mellow
Strong	Sensitive	Lively	Cheerful	
6. Light	**7. Blue**	**8. Temperature**	**9. Purity**	
Luminous	Tranquil	Warm	Intricate	
Misty	Calm	Subtle	Bold	
Deep	Soothing	Cool	Innocent	
		Cold	Pure	

3.4 Implementation

The scope of this research was to develop a prototype system using a local database of images, but implement it using technologies that could be easily adapted for an online environment. With this in mind, the Flickr photograph collection was chosen in order to build a local cache of images. This section describes the implementation of X2Photo.

3.4.1 Data Collection

This research's experimental approach called for sets of arbitrary photographs to be collected with no distinct styles. Therefore, photographs needed to be cached from a large number of users. The ideal way to realise this requirement was to query Flickr for a list of public photos. Using this approach, more than 12,000 random photographs from Flickr were cached in small, medium, and large sizes. The Flickr API was also used to retrieve the metadata related to the cached photographs. Once the necessary data for each photograph was parsed and stored, a tags repository was created based on unique tags within the collection. A list of tags related to the given tag, based on clustered usage analysis within Flickr, was also stored. At this stage it was realised that the metadata from Flickr alone was not extensive enough to capture an image's aesthetics. Only objective concepts such as when and where the photograph was taken and whether it was an indoor or outdoor image could be derived.

To create an aesthetic vocabulary that supported colour theory required more metadata about the images to be obtained. This meant that each digital photograph's pixel values were analysed in order to get its dominant tones and colours. Red, green and blue values for each pixel were then extracted and within each block these values were rounded to their contextual RGB values to avoid almost duplicate colours. A

photograph's hue had to be processed with a finer quantisation as human perception is more sensitive to hue than saturation and lightness. Once an image was processed all its data was combined into a uniform model. This schema was then registered with SARA so that semantic attributes could be formed from them, and so that this media repository could be linked to other ones if so desired.

3.4.2 User Interface and Architecture

Figure 1 shows the front-end of the system in which the three main areas of the interface can be seen. The main part of the screen is called the Discovery Space and contains the result set of photographs from a user query. The wall of photographs can be dragged by user and individual images selected to see what tags and semantic attributes are associated with it.

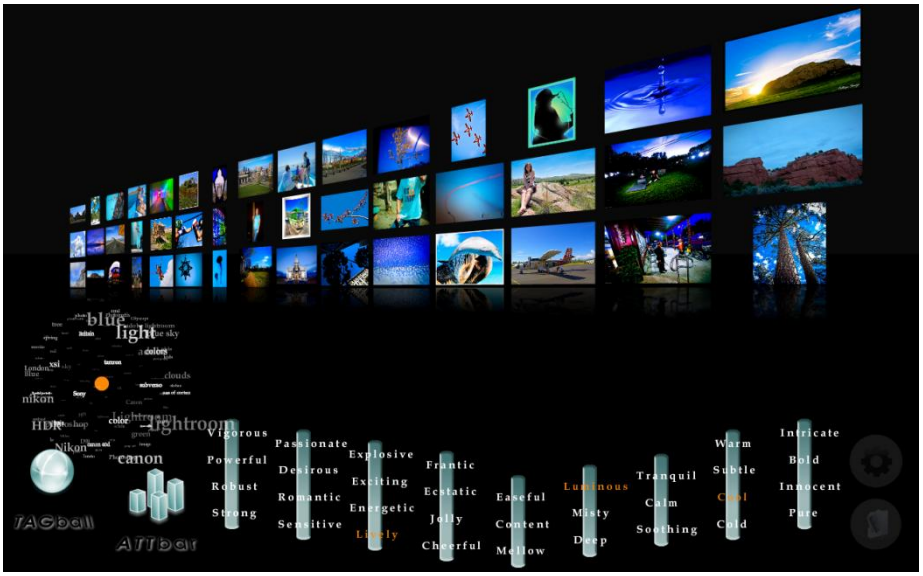

Fig. 1. The X2Photo Interface

The bottom of the screen is dominated by the AttBar which represents each of the nine semantic attributes as a vertical bar. Each bar contains the different parameters relating to each semantic attribute, with Figure 1 showing the results from a query containing the aesthetics *Lively, Luminous* and *Cool* from the AttBar. The user can select a parameter from each bar to add a query. Using SARA's facility for tailoring of domain expertise, it would also possible for end-users not happy with the results they were getting to alter the rules associated with each semantic attribute, so that a different range of images were returned for that particular parameter.

The single dimensional nature of tags has already been discussed, highlighting that they are limited in communicating the aesthetic values of photographs, but are instead more useful in defining their content. Thus, in order to help the user find an image with specific content, the system had to show any tags associated with the result collection of photographs, as well as those from each individual photograph. By integrating this with support for aesthetic exploration of images, it gives users a more

flexible way of finding relevant photographs. The number of tags typically exceeds a number that could be clearly displayed with a simple tag clouds, hence a TagBall was used instead to allow large numbers of tags to be displayed while not cluttering the UI. In the bottom left hand corner of Figure 1 is the TagBall which displays all the Flickr tags related to the entire result set or individual photograph. The user just has to select any of these to refine their searches.

When users click on an image, the interface zooms into the photograph, and they can click the flip button on the top left corner of the photograph to see its details. If a user clicks the down arrow icon, the photograph is brought into focus, with the AttBar displaying its associated semantic attributes, and the TagBall displaying the relevant tags. A user can store the photograph to the Favourites area by clicking the star button. Figure 2 shows the overall architecture of the application. A user forms queries by selecting semantic attributes in the X2Photo GUI, with all queries sent to SARA via its parameter based API. Because the metadata relating to the images are stored as XML, XQuerys encapsulated within each semantic attribute are sent to the database storing the metadata and the relevant image identifiers returned for each semantic attribute. These separate result sets (if the query contains more than one semantic attribute) are then reconciled into a consolidated set which is returned to X2Photo in XML. The corresponding images from the photograph cache are then rendered to the user in X2Photo's GUI.

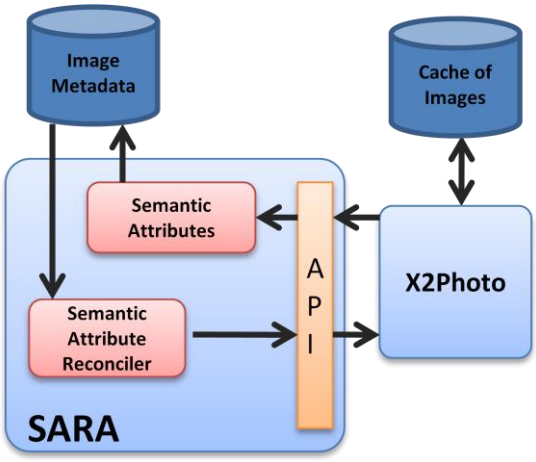

Fig. 2. X2Photo Architecture

4 Evaluation

X2Photo was evaluated to test the usability, functionality and the overall appeal of the system. Furthermore, the potential benefits of injecting subjective expert knowledge based on the manipulation of non-textual, low-level data, in comparison to conventional methods (like image retrieval via tags only) was also examined. With this aim in mind the following approach was pursued: Four photographs (see Figure 3) not present in the 12,000 random photographs collection were selected. These photographs were then prescribed different semantic attributes by the domain expert,

who was an experienced photographer. The nine users were first shown all four photographs, and were then asked to freely describe them in their own words. Then they were given an overview of the tool and were asked to do the following tasks:

- For photograph 1, find similar photographs via X2Photo
- For each photograph found, add it to the Favourites.
- Repeat this task for all four photographs.
- Once complete, go to Flickr and for photograph 1; again try to find similar images either with the words originally used to describe the photographs or with different ones.
- Repeat this task for all four photographs.

After finishing these tasks they were given a survey to fill out, to complete the user-test.

(a) **(c)**

(b) **(d)**

Fig. 3. The four initial photographs shown to users

4.1 Describing the Images

The majority of users evaluating X2Photo were technically proficient with computers, and four considered themselves to be amateur photographers. How the different users described the four photographs had some noteworthy aspects, such as those users who were interested in photography tending to use more technical phrases. For instance,

some wrote terms such as “over-exposed” when describing the photograph *a*, mentioned the angle at which the photograph *b* might have been shot at, and questioned whether this photograph was altered in an image editing program to obtain its deep contrast. These users tended not to describe the content of the photograph as much as the users with little photography experience. Some users preferred to describe the photographs with more personal expressions such as “lonely” and “tempting” when referring to photograph *c*. Photograph *d*, as expected, was interpreted differently by almost all the users. While some tried to figure out what the man in the picture might be doing, some chose to describe him, resulting in many different impressions such as “gritty”, “relaxed” or “run-down”.

Almost all the users first chose expressions like “warm”, “cold”, “airy”, “gloomy” and “energetic”, some of which directly coincided with the actual attributes determined by the domain expert. They then proceeded to describe the actual content. Two of the nine participants were more objective in their descriptions and chose to name the elements they saw in the photographs with words like “corridor”, “bench”, “rocks”, “back alley”, etc. However, the vast majority of users combined their perceptions with the content: “...a cool calm picture but alive…there's a woman sitting on a bench... feels breezy but soft... waves look relaxing”.

4.2 Finding Images in X2Photo

Just like their preferences in describing the photographs, the users’ approach to finding similar photographs in X2Photo were particularly different. Four users never actually enabled the TagBall. Coincidentally their descriptions of the photographs were heavily consisted of expressions like “moody”, “dark”, “calm”, etc. They directly chose similar words present within the AttBar and then carried out their searches. After receiving their initial results two of these users were surprised to see how the tool interpreted their descriptions. They did not agree with the expert and started experimenting with the AttBar rather than continuing with their searches. After observing some consecutive result sets and bringing some photographs into focus, they stated that they grasped the association the expert was making, and modified their searches accordingly. The other two users who didn’t use the tag ball performed 2-3 consecutive searches which were refined each time, to find a similar photograph. Observing the similar photographs that users returned, it was interesting to see what the users based their similarity criteria on. While some photographs have a similar feel to them regarding the concept or the context, some are similar in content as well. Figure 4 shows examples of similar pictures (of images b and c in Figure 3) found by users using the TagBall and AttBar in X2Photo

4.3 Finding Images in Flickr

When the users tried to find similar photographs in Flickr, their approaches were again different. For example, one user used “fiery clinical harsh” to search for the photograph a, which were the expressions he had used when describing the photographs originally. In contrast, another abandoned their expressive vocabulary used originally to describe photograph b (because he was very familiar with searching on Flickr) and chose to use the search phrase “Scotland cliff coast”. Some users were very articulate in their searches and submitted phrases such as “city lights low angle

journalistic lonely" to find similar photographs. With three user's searches within Flickr, a slight change in their vocabulary could be seen. For instance, a user who had previously described the photograph c mainly based on content; "beach, person sitting on the bench, greyish" found a similar image within X2Photo that the expert thought to be "romantic", "soothing" and "innocent". Within Flickr, the user thus carried out his first search with the terms "romantic sea scenery". Users familiar with Flickr also used the advanced search available and refined their queries, but again tended to use content-based terms to carry out their searches. In the end all the users were able to find at least one similar image, which was not surprising considering the amount of photographs Flickr has. However it was noteworthy how all the users had to resort to content-based terms (identical in many cases). This showed how such systems can limit the ways individuals search for photographs.

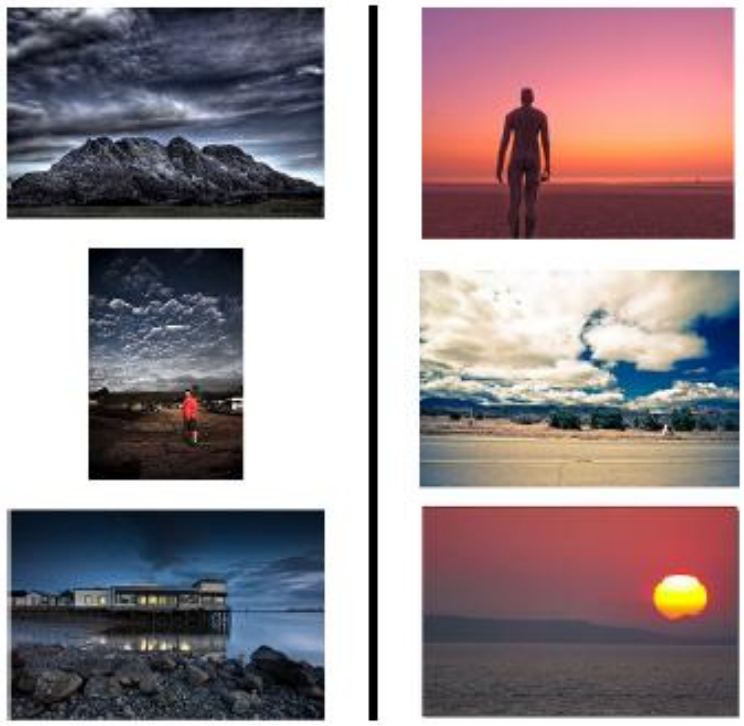

Image (b) **Image (c)**

Fig. 4. Examples of similar pictures found by users using the TagBall and AttBar in X2Photo

4.4 User Survey

A user survey was also conducted once the given tasks were completed. The questionnaire intended to evaluate each feature's functionality as well as aesthetic qualities, and also the overall system quality regarding various aspects. The general response to the usability and appeal of the Discovery Space was very positive,

agreeing that the continuous flow enabled them to browse the photographs thoroughly, and that the interaction with the space was appealing. Below is a summary of the survey results:

- 8/9 users considered the overall UI to be very good
- 8/9 strongly agreed that the system was attractive
- 8/9 users found the zoom effect in the interface to be very good or good
- 9/9 users thought the AttBar was very good or good, that the concept was comprehensible and the classification of the attributes were clear.
- 8/9 users found the ability to refine a search with a focus image to be very useful or useful. By seeing what semantic attributes and tags were associated with the focus image it allowed them to use these as a springboard for their browsing.

4.5 Analysis

The user experiment and the survey that followed suggested that when describing photographs, whether interested in photography or not, people like to communicate "how" a photograph is as well as "what" it portrays. This finding indicates a need for a wider vocabulary to be available to users in order to retrieve accurate and relevant photographs from any collection. Traditional tag-based systems tend to be dominated by content-based terms, thus ignoring the artistic quality which is a key factor that evokes appreciative emotions. Hence these systems often reduce photographs to a list of mainly content-based words. As most people have become accustomed to this approach, in such an environment they tend to ignore other ways in which they would approach a photograph, and are therefore relegated to search for the tagged simplification of a photograph, rather than the actual photograph itself.

Based on the photographs found by the users when using the more natural expressions via the X2Photo system, it indicated that this approach could grant users the additional useful axes to when searching for photographs. Thus, injecting expert knowledge, based on the manipulation of raw low-level data, into a conventional system only supporting tag-based search, allows users to more freely express both the photograph and the picture it is conveying. Even though a specific expert vocabulary may not be suitable or correct for each individual, users can adapt to the expert's view or better yet choose to subscribe altogether to a different expert expanding the semantic space. SARA also provides the functionality for end users to tailor an expert's semantic attribute to better fit their own vocabulary.

X2Photo received overall positive feedback; the users clearly understood the idea and the overall concept. They suggested that users should be able to subscribe to different experts and that there should be a more comprehensive range of semantic attributes. This all indicates that the users understood the aim of the tool and how it could be extended further. All users agreed that they could see a real-life application of the tool if some further improvements were made and they could see themselves utilising such a tool in their everyday lives. In order to offer users an alternative way of finding a photograph, a system has to have a rich vocabulary. Hence, by increasing the range of low level features, it would enable experts to create more refined semantic attributes, resulting in a more useful system for end users.

5 Summary

This paper investigated the possible benefits of augmenting the conventional tag-based query techniques used in many image databases, with subjective expert knowledge built on raw low-level data. Based on this notion, X2Photo was developed which aimed to empower users in retrieving photographs from collections, using not only objective tags from Flickr, but also subjective expertise based on a photograph's colour space. Semantic attributes were encoded into the system via SABer, which provided end users with semantically meaningful access points into the domain, which were not previously available.

The user test and the results of its accompanying survey, highlight how people like to communicate the aesthetics of a photograph as well as what it portrays. However, when utilising a conventional tag-based system, they tend to ignore the aesthetics and emotions conveyed in the images, as the tag-based systems tend to be overloaded towards content based tags. Hence this can lead to limited searching via tagged simplifications of a photograph, ignoring the aesthetics of the photograph. The types of photographs found by the users, using the more natural expressions offered by the system, indicate that this approach can be used to grant users more versatility when searching for photographs. Hence, injecting expert knowledge into a conventional system that only offers tag-based searching, would allow users to freely express both the aesthetic of the photograph they want, as well as the picture it conveys. This offers users an alternative pathway to access large photograph collections.

All the users agreed that the system was a powerful tool for exploring photographs and when asked if they could see a real-world application stemming from X2Photo, all users concurred, as long as further improvements were made. Some engineering decisions need to be reconsidered in order to offer a more robust system and an alternative approach to manipulating the tags would be beneficial. The number of semantic attributes could also be increased, and the extraction of the underlying features improved with more sophisticated image analysis techniques. This would enable experts to create more refined semantic attributes. Moreover as a specific expert vocabulary may not be suitable or correct for each individual, users should be able to subscribe to different experts to consider different perspectives.

Leading image search engine such as Google Images have recently provided a few colours to be selected in order to have results with similar colour spaces. Considering the huge volume of images they index, and this new functionality they offer, it can be suggested that the methodology proposed in this paper could be integrated seamlessly into online image searching. This new functionality would allow users to pose verbal queries rather than selecting some basic colours to match. This approach could be also applied to other media such as video and audio, with SARA supporting experts in those fields to classify characteristics that end users could leverage in their searches. Likewise, multiple experts from the same domain can be supported by SARA, with the end user able to select characteristics created by different experts and tailoring them to their own needs if necessary. Finally, because SARA provides a consolidated interface to multiple sources from a domain, it can support applications that give users powerful searching and browsing operations over many separate image repositories.

Acknowledgments. This research has been supported by The Irish Research Council for Science, Engineering and Technology: funded by the National Development Plan.

References

1. Cui J, Wen F, Tang X, Real time google and live image search re-ranking. In: Proceeding of the 16th ACM international conference on Multimedia, pp. 729-732 (2008)
2. Jackson R.S (2002) Wine tasting: a professional handbook. Elsevier
3. Hampson C, Conlan O, Leveraging Domain Expertise to Support Complex, Personalized and Semantically Meaningful Queries Across Separate Data Sources. In: Proceeding of the Fourth IEEE International Conference on Semantic Computing (ICSC'10), Pittsburgh, USA, pp. 305-308 (2010)
4. Hare J.S, Lewis P.H, Enser P.G.B, Sandom C.J, Mind the gap: Another look at the problem of the semantic gap in image retrieval. In: Multimedia Content Analysis, Management, and Retrieval 2006. 6073, pp.75-86 (2006)
5. Enser P.G.B, Sandom C.J, Lewis P.H, Surveying the reality of semantic image retrieval. In: LNCS, vol. 3736, pp. 177-188, Springer, Berlin / Heidelberg (2006)
6. Datta R, Li J, Wang J.Z, Content-based image retrieval: approaches and trends of the new age, In Proceedings of the 7th ACM SIGMM international workshop on Multimedia information retrieval pp. 253-262. ACM, New York (2005)
7. Tamura H, Yokoya N, Image database systems: A survey. Pattern Recognition. 17, pp. 29-43 (1984)
8. Shen H.T, Ooi B.C., Tan, K.L, Giving meanings to WWW images. In: Proceedings of the 8th ACM international conference on Multimedia, pp. 39-47. ACM, New York (2000)
9. Cai D, He X, Li Z, Ma W.Y, Wen, J.R, Hierarchical clustering of WWW image search results using visual, textual and link information, In: Proceedings of the 12th annual ACM international conference on Multimedia, pp. 952-959. ACM, New York (2004)
10. Marlow C, Naaman M, Boyd D, Davis M, Position paper, tagging, taxonomy, flickr, article, to read. In: Collaborative Web Tagging Workshop, Edinburgh, Scotland (2006)
11. Gong Y, Advancing content-based image retrieval by exploiting image color and region features. Multimedia Systems. 7, pp.449-457 ACM, New York (1999)
12. Yu H, Li M, Zhang H.J, Feng, J, Color texture moments for content-based image retrieval. In: Proceedings of the International Conference on Image Processing, pp. 929-932 (2002)
13. Shih J.L, Chen L.H, Color image retrieval based on primitives of color moments. In: LNCS, vol. 2314, pp. 88-94, Springer, Berlin / Heidelberg (2002)
14. Davis S, Color perception: Philosophical, Psychological, Artistic, and Computational Perspectives. Oxford University Press (2000)
15. Gage J, Color and Meaning: Art, Science, and Symbolism. University of California Press, Berkeley (1999)
16. Fehrman K, Fehrman,C.F, Color: The Secret Influence, Prentice Hall (2000)
17. Valdez P, Mehrabian A, Effects of color on emotions. Journal of Experimental Psychology. 123, pp.394-408 (1994)
18. Parramon J, Color Theory. Watson-Guptill Publications, New York (1989)
19. Sinha P, Jain R, Semantics In Digital Photos A Contextual Analysis. In: Proceeding of the Second IEEE International Conference on Semantic Computing, pp. 58-65 (2008)
20. Enser P, Visual image retrieval: seeking the alliance of concept-based and content-based paradigms. Journal of Information Science. 26, pp.199-210 (2000)